ACCURATE SVDs OF
– Weakly Diagonally Dominant M-Matrices
– Polynomial Vandermonde Matrices

Plamen Koev
Department of Mathematics
University of California - Berkeley

Joint work with James Demmel

Supported by NSF and DOE

International Workshop on Accurate Solutions of Eigenvalue Problems
Split, Croatia, June 24-27, 2002

Accurate SVDs of Weakly Diagonally Dominant M-Matrices

- Def: M-Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix};$$

$a_{ii} \geq 0$

$a_{ij} \leq 0, \; i \neq j$

Row Sums $s_i = \Sigma_{j=1}^{n} a_{ij} \geq 0$

- Given: Row sums s_i and off diagonals $a_{ij}, i \neq j$.
- Goal: Compute the SVD of A to high relative accuracy:
  - $|\sigma_i - \hat{\sigma}_i| \leq O(\epsilon)\sigma_i$
  - $\theta(w_i, \hat{w}_i) \leq O(\epsilon)/\min_{i \neq j} \frac{|\sigma_j - \sigma_i|}{\sigma_i}$
- Diagonal elements computable accurately, sum of positives

$$a_{ii} = s_i - \sum_{j \neq i} a_{ij}$$

- Suffices to perform GECP accurately
- Then use SVD algorithm of Demmel, Gu, Eisenstat, Slapničar, Veselić, Drmač

GECP on Weakly Diagonally Dominant M-Matrices

- **Pivoting is diagonal, preserves structure**
- **One step of GE:**
  - **Off diagonals:** $a_{ij} = a_{ij} - \frac{a_{ik}a_{kj}}{a_{kk}}$
  - **Row sums:** $s_i = s_i - \frac{a_{ik}}{a_{kk}} s_k$
- **Everything is preserved in Schur complementation**
  - **Weak diagonal dominance**
  - **M-matrix structure**
  - **High relative accuracy in** a_{ij} **and** s_i
- **Yields LDU from GECP thus SVD to high relative accuracy componentwise**
- **Small changes in** $a_{ij}, i \neq j$ **and** s_i **only cause small relative changes in the SVD**
- **NO ILL CONDITIONED SET (no relative gaps to worry about)**

SVDs of Polynomial Vandermondes – Definitions and Goals

- P – basis of polynomials, $\deg P_i = i - 1$

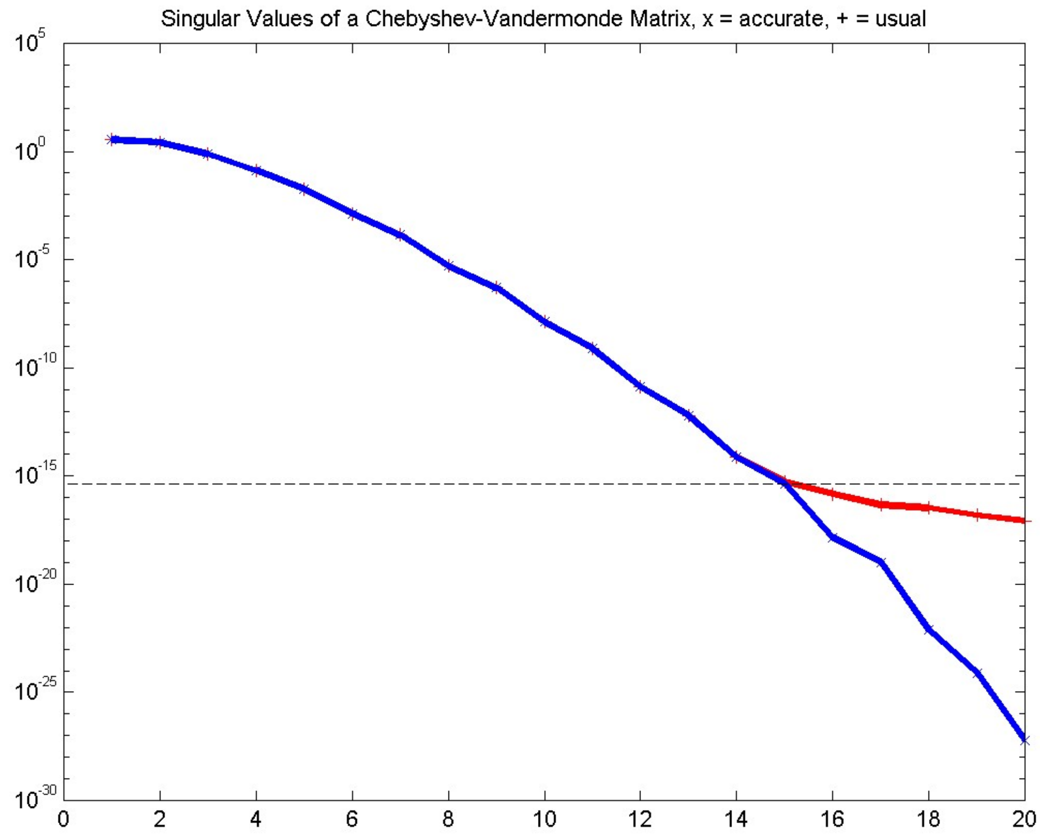
$$V(x) = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \dots & P_{n-1}(x_1) \\ P_0(x_2) & P_1(x_2) & \dots & P_{n-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \dots & P_{n-1}(x_n) \end{bmatrix},$$

  is called ***Polynomial Vandermonde*** matrix.

- Of particular interest: orthonormal polynomials, monomials
- $V(x)$ – Notoriously ill-conditioned
- Goal: Compute the SVD of $V(x) = W \cdot \Sigma \cdot Z$ to high relative accuracy, regardless:
  - $|\sigma_i - \hat{\sigma}_i| \le O(\epsilon)\sigma_i$
  - $\theta(w_i, \hat{w}_i) \le O(\epsilon)/\min_{i \neq j} \frac{|\sigma_j - \sigma_i|}{\sigma_i}$
- Motivation
  - SVD determined accurately by the data
  - Right SVD costs the same as the wrong SVD – $O(n^3)$

Example: 20 by 20 Chebyshev-Vandermonde Matrix

- $V_{ij} = T_{i-1}(x_j)$, $x = \text{random} \in [0, 0.2]$; σ_i range from 10^0 down to 10^{-29}
- Old algorithm, New Algorithm, both in 16 digits



- $D = \log(\text{cond}(A)) = \log(\sigma_1/\sigma_n)$ (here $D = 28$)
- Cost: Old algorithm $= O(n^3 D^2)$, New algorithm $= O(n^3)$, independent of D

OVERVIEW

- **Orthogonal and orthonormal polynomials**
- **Model of Arithmetic:** $\mathbf{1+\delta}$
- **Demmel 1998: SVD of Cauchy matrix to high relative accuracy**
- $\boldsymbol{V(x) = C(x,y) \cdot V(y)}$**, where** $\boldsymbol{C(x,y)}$ **– Cauchy**
- **Can choose** $\boldsymbol{y}$**:** $\boldsymbol{V(y)} = \textbf{DIAGONAL} \cdot \textbf{ORTHOGONAL}$
- $\boldsymbol{V(x)} = \underbrace{(\mathrm{C(x,y)} \cdot \textbf{DIAGONAL})}_{\textbf{CAUCHY}} \cdot \textbf{ORTHOGONAL}$
- **Analytical** $\boldsymbol{y}$ **vs. numerical** $\boldsymbol{\hat{y}}$

ORTHOGONAL AND ORTHONORMAL POLYNOMIALS

- Orthogonal Polynomials: deg$P_n = n - 1$, orthogonality relation

$$\int_a^b w(x)P_m(x)P_n(x)dx = \delta_{mn}c_n,$$

$w(x)$ – weighting function. $c_n = \int_a^b w(x)P_n^2(x)dx$

- If $c_n = 1$ *orthonormal.* Can be normalized easily: $\frac{P_n(x)}{\sqrt{c_n}}$
- Useful in the solution of mathematical and physical problems.
- Provide a natural way to solve, expand, and interpret solutions to many types of important differential equations.
- Easy to generate using Gram-Schmidt orthonormalization.
- Roots: Simple, real, in $[a, b]$. Roots of P_i interlace those of P_{i+1}
- Satisfy three-term recurrence

$$P_{n+1}(x) = (d_n x + b_n)P_n(x) - a_n P_{n-1}(x), \qquad P_0(x) = const, P_1(x) = d_0 x + b_0$$

- *Discrete orthogonality property*: If ξ_i are the roots of P_n, then:

$$\sum_{i=1}^{n} d_i P_m(\xi_i)P_k(\xi_i) = \delta_{mk}, \text{ for } k, m < n, \qquad d_i - \text{Christoffel numbers}$$

SOME FAMOUS ORTHOGONAL POLYNOMIALS

Type	Interval	$w(x)$	c_n
Chebyshev, 1st kind	$[-1, 1]$	$(1 - x^2)^{-1/2}$	$\pi, \left(\frac{\pi}{2} \text{ for } n = 0\right)$
Chebyshev, 2nd kind	$[-1, 1]$	$\sqrt{1 - x^2}$	$\frac{\pi}{2}$
Hermite	$(-\infty, \infty)$	e^{-x^2}	$\sqrt{\pi} 2^n n!$
Laguerre	$[0, \infty)$	e^{-x}	1
Legendre	$[-1, 1]$	1	$\frac{2}{2n+1}$

Model of Arithmetic: $1 + \delta$

- $\mathrm{fl}(a \otimes b) = (a \otimes b)(1 + \delta)$, no under/overflow
- How can we lose accuracy in computing in floating point?
  - OK to multiply, divide, add positive numbers
    Proof: $1 + \delta$ factors can be factored out
  - $x_i \pm x_j$, where x_i and x_j are initial data (so exact)
  - Cancellation when subtracting approximate results dangerous:

```
   .12345xxx
-  .12345yyy
------------
   .00000zzz
```

- We will compute everything using only allowable expressions

Demmel 1998: SVD of a Cauchy Matrix to High Relative Accuracy

- Cauchy: $C(i,j) = \frac{1}{x_i + y_j}$.
- Phase 1: $C = LDU$ from GECP via (implicitly) computing accurate minors
- Phase 2: Compute the SVD from $C = LDU$

How do we compute $A = L \cdot D \cdot U$ for a Cauchy Matrix?

- Cauchy:

$$C(i,j) = \frac{1}{x_i + y_j}, \qquad \det(C) = \frac{\Pi_{i<j}(x_j - x_i)(y_j - y_i)}{\Pi_{i,j}(x_i + y_j)}$$

  – No bad cancellation ⇒ good to most digits

- Change inner loop of Gaussian Elimination from

$$C(i,j) = C(i,j) - \frac{C(i,k)C(k,j)}{C(k,k)}$$

  to

$$C(i,j) = C(i,j)\frac{(x_i - x_k)(y_j - y_k)}{(x_k + y_j)(x_i + y_k)}$$

- Each entry of L, D, U accurate to most digits!
- (Quasi-)Cauchy:

$$\text{diag}(a_1, ..., a_n) \cdot C \cdot \text{diag}(f_1, ..., f_n) = \left[\frac{a_i f_j}{x_i + y_j}\right]_{i,j=1}^{n}$$

  – analogous

The Vandermonde – Cauchy Connection

- Let $x = (x_1, ..., x_n), \quad y = (y_1, ..., y_n), \quad x_j \neq y_j$

$$a_i = \prod_{j=1}^{n} (x_i - y_j), \qquad f_i^{-1} = \prod_{j=1, j \neq i}^{n} (y_i - y_j)$$

  – computable accurately

- Thm (V. Olshevsky, 1995):

$$V(x) = C(x, y) \cdot V(y)$$

  where $C(x, y) = \left[\frac{a_i f_j}{x_i - y_j}\right]_{i,j=1}^{n}$ – CAUCHY

- SVD of $C(x, y)$ – OK
- Suffices to find y such that $V(y)$ is orthogonal

How do we choose y in $V(x) = C(x,y) \cdot V(y)$

- ***Discrete Orthogonality Property***:

$$\sum_{i=1}^{n} d_i P_r(y_i) P_s(y_i) = \delta_{rs}$$

  where

  – y – roots of P_n

  – d_i – Christoffel numbers

- $DV(y) = Q$ – orthogonal, where $D = \mathrm{diag}(\sqrt{d_i})_{i=1}^{n}$
- $V(x) = \underbrace{C(x,y) \cdot D^{-1}}_{\text{CAUCHY}} \cdot \underbrace{D \cdot V(y)}_{Q} = \underbrace{W \cdot \Sigma \cdot Z}_{\text{DEMMEL '98}} \cdot Q$

Computing y = roots of P_n, and d_i

- **Not a part of the algorithm, needs to be done once for all $V(x)$**
- **Roots with small relative error – OK**
- **y_i – eigenvalues of a tridiagonal matrix (from the three term recurrence)**
- **$d_i = \det(A(1:i, 1:i))$ – also accurate**
- **Extremely well conditioned for Chebyshev, Legendre, Laguerre, Hermite**
- **For Chebyshev, 2nd kind:**

$$A = \begin{bmatrix} 0 & .5 & & & & \\ .5 & 0 & .5 & & & \\ & .5 & 0 & .5 & & \\ & & \ddots & \ddots & \ddots & \\ & & & .5 & 0 & .5 \\ & & & & .5 & 0 \end{bmatrix}$$

The case $x_i = y_j$ in $C(x, y)$

- Removable singularity
- 2×2 example

$$\begin{bmatrix} (x_1 - y_1)(x_1 - y_2) & \\ & (x_2 - y_1)(x_2 - y_2) \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{x_1 - y_1} & \frac{1}{x_1 - y_2} \\ \frac{1}{x_2 - y_1} & \frac{1}{x_2 - y_2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{y_1 - y_2} & \\ & \frac{1}{y_2 - y_1} \end{bmatrix}$$

$$= \frac{1}{y_2 - y_1} \begin{bmatrix} x_1 - y_2 & x_1 - y_1 \\ y_2 - x_2 & y_1 - x_2 \end{bmatrix}$$

The case $x_i = y_j$ in $C(x, y)$

- Removable singularity
- 2×2 example

$$\begin{bmatrix} (x_1 - y_1)(x_1 - y_2) & \\ & (x_2 - y_1)(x_2 - y_2) \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{x_1 - y_1} & \frac{1}{x_1 - y_2} \\ \frac{1}{x_2 - y_1} & \frac{1}{x_2 - y_2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{y_1 - y_2} & \\ & \frac{1}{y_2 - y_1} \end{bmatrix}$$

$$= \frac{1}{y_2 - y_1} \begin{bmatrix} x_1 - y_2 & x_1 - y_1 \\ y_2 - x_2 & y_1 - x_2 \end{bmatrix}$$

- Inner loop of Gaussian Elimination

$$C(i, j) = C(i, j) \frac{(x_i - x_k)(y_j - y_k)}{(x_k + y_j)(x_i + y_k)}$$

still valid if $C(i, j) \neq 0$

- If $C(i, j) = 0$, do the "obvious":

$$C(i, j) = C(i, j) - \frac{C(i, k)C(k, j)}{C(k, k)} = -\frac{C(i, k)C(k, j)}{C(k, k)}$$

still accurate.

Any Polynomial Vandermonde

- OK, if we can find y such that

$$V(y) = \text{DIAGONAL} \cdot (\text{WELL CONDITIONED}).$$

- Perfect Example: Ordinary Vandermonde (Demmel 1998)

$$V(x) = \left[x_{i-1}^{j}\right]_{i,j=1}^{n}; \qquad V\left(\left[e^{\frac{2i\pi}{n}}\right]_{i=1}^{n}\right) = \text{DFT}$$

- Unnormalized orthogonal polynomials?
  OK, if the normalization factors don't vary too much

$$V(x) \cdot F = C(x,y)V(x) \cdot F = C(x,y) \cdot D \cdot Q \cdot F$$

  still polynomial complexity for Chebyshev, Laguerre, Hermite, Legendre.

- Open problem:
  Confluent Polynomial Vandermonde, we think No.

- These slides: `www.math.berkeley.edu/~plamen/iwasep4.ps`
- My Ph.D. thesis, `www.math.berkeley.edu/~plamen/a.ps`

- Totally positive linear systems can be solved extremely accurately if minors of matrices factor e.g. Vandermonde, Cauchy
  - Björck-Pereyra-type algorithms
  - New result: Generalized Vandermonde matrices
  - New faster algorithm for computing the Schur function
- Computer Arithmetic
  - Impact of model of arithmetic on algorithms
    e.g. $(1 + \delta)$ vs. IEEE
  - Harry Diamond's Theorem and the Table Maker's Dilemma
    f
    $g = f^{-1}$
    $H = G \circ F$
    $H(H(H(x))) = H(H(x))$
  - Proved validity when f non strictly convex and IEEE arithmetic rounding to nearest even
- Computational noncommutative algebra
  - Polynomial Identities in Matrix Algebras
  - All identities of degree $2k + 2$ in $M_k(\mathbf{Z})$ follow from the
    Standard Identity of Amitzur-Levitzky of degree $2k$.
- http://www.math.berkeley.edu/~plamen/a.ps